Binaural sound source localization in real and virtual rooms

Monika Rychtáriková[1,3], Tim van den Bogaert[2], Gerrit Vermeir[1], Jan Wouters[2]

[1]*K.U. Leuven, Laboratory of Acoustics and Thermal Physics, Celestijnenlaan 200D, bus 2416, 3001 Heverlee, Belgium*

[2]*K.U. Leuven, Exp ORL, Dept. Neurosciences, O&N2, Herestraat 49, bus 721, 3000 Leuven, Belgium*

[3]*STU Bratislava, Faculty of Civil Engineering, Dept. of Building Constructions, Radlinského 11, 81108 Bratislava, Slovakia*

Email address: Monika.Rychtarikova@bwk.kuleuven.be

Submitted October 26, 2008

Revised February 10, 2009

ABSTRACT

On the basis of a statistical evaluation of the binaural sound source localization performance during listening tests by human subjects, we show that the convolution of a measured head-related transfer function (HRTF) with the room impulse response generated by a hybrid image source model with a stochastic scattering process using secondary sources provides an adequate model for the prediction of binaural room impulse responses (BRIR) for directional sound localization in frontal horizontal plane.

The source localization performance for sound stimuli presented in a natural way and via headphones to test subjects was compared. Listening via headphones tends to decrease the localization performance only and just slightly when localizing narrow-band high-frequency stimuli. Binaural sound presented to test subjects via headphones was generated on the basis of measurements and simulations.

Two different headphone conditions were evaluated. The overall localization performance for simulated headphone sound obtained using a simulated BRIR was found to be equivalent with the one using a measured BRIR.

The study also confirms expectations that the deterioration of sound source localization performance in reverberant rooms are closely linked with the direct-to-reverberant ratio for given speaker and listener positions, rather than with the reverberation time of the room as such.

PACS : 43.55.Ka, 43.66.Qp , 43.66.Pn

0 INTRODUCTION

In recent years intensive research has resulted in the development and implementation of powerful noise suppression algorithms in modern digital auditory prostheses: hearing aids (HA) and cochlear implants (CI). Good results have been obtained by different approaches, using multi-microphone adaptive filter techniques that rely on spectral as well as spatial information [1-4]. However, these techniques are typically developed and evaluated in monaural systems. These modern adaptive signal processing approaches are not optimized for the preservation of interaural information when used bilaterally. A correct presentation of interaural differences (in amplitude and time) is a prerequisite for localizing sound sources and for improved speech understanding noise due to the spatial release from masking and is often not taken into account. Many of the speech processing algorithms often work well in laboratory conditions and in rooms with short reverberation times, but in real-life acoustical situations they may strikingly fail.

Most studies focus on monaural speech understanding issues in different acoustical environments [8-10] and presently different studies are being carried out towards modeling binaural speech intelligibility. Different speech intelligibility models based on the combination of the equalization-cancellation model with binaural extensions of the speech intelligibility index or the speech transmission index, are nowadays enhanced to incorporate binaural listening situations under various conditions [11,12]. In this research, measured and simulated conditions are used, and listening tests were made with both normal and hearing impaired subjects. Although simulation software (such as ODEON® [13]) has often been used to model the acoustical environments, perceptual validation for speech intelligibility purposes was done only very recently [14] Furthermore, no in-depth validation of simulation software has ever been carried in terms of predicting the impact of the virtual acoustical environment

on directional hearing or localization of sound. In our research we therefore deal with this important aspect of the virtual models of rooms.

Recent studies have shown that signal processing algorithms for noise reduction and compression, as implemented in modern commercial hearing aids, can lead to distortions and have a large impact on the localization performance of hearing aid users [6,7,15]. This illustrates the need to validate signal processing algorithms with respect to their preservation of binaural cues and spatial awareness.

Performing perceptual evaluations with normal hearing and hearing impaired subjects using sound localization test set-ups in different rooms with different acoustical properties (such as churches, anechoic rooms, restaurants, halls or railway stations) is very laborious and time consuming, and therefore very difficult to realize in practice. Virtual acoustics may be very useful for the evaluation and development of signal processing strategies for auditory prostheses. However, the question remains to which extent these virtual acoustic techniques result in an accurate prediction of real-life performance of the different algorithms.

Different methods have been used in the past to create virtual auditory environments. Two approaches are envisaged to reproduce real-life acoustical environments: on the one hand the presentation through multi-loudspeaker arrays of multi-channel recordings, and on the other hand acoustical modeling of real rooms (listening environments) combined with binaural presentation to the listener via headphones or direct audio inputs.

Though such systems may successfully generate perceptions of spatial hearing, it is not straightforward to satisfy the quite stringent accuracy requirements concerning binaural cues. The Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) have to be reproduced accurately enough for correct spatial hearing.

A virtual auditory space can be synthesized using a so-called cross-talk cancellation system [16-22]. Cross-talk cancellation systems use digital filters based on measured head-related impulse responses (HRIR). The generation of the final cancellation is based on these HRIRs. However, several research studies have revealed some limitations of this technique. Several authors [23-25] have shown that many of the systems introduce front-back confusions, particularly if only two loudspeakers are used, and large errors appeared close to the target positions at ±90°. Several studies also investigated binaural performance of a multiple-loudspeaker system [22,26]. With these binaural cross-talk systems difficulties have been reported for delivering accurate ITD and ILD cues.

Another approach for recreating the spatial soundscape with multi-channel loudspeaker setups is the spatial impulse response rendering method. It has been shown to be successful to some degree in creating virtual acoustical environments [27,28]. A major drawback of most of these methods is that they have not (or only in a minor way) been evaluated and validated perceptually. The latter has to focus on directional hearing (sound localization) and speech understanding (spatial unmasking) which are of extreme importance in daily life, both for normal and impaired hearing listeners.

Other possibilities of acoustical reconstruction of real-life auditory situations in laboratory or clinical conditions are through headphone listening tests. These experiments can be based on binaural impulse response measurements *in situ*, e.g. using an artificial head, or by advanced acoustic modeling of a room in prediction software [14,29]. The latter represents a very flexible approach for preparing listening tests in terms of easily replacing head-related transfer functions (individual or artificial head) and in terms of swiftly changing the acoustics of a virtual room without having to measure impulse responses in this room. Head-related transfer functions represent the impulse response of an incoming quasi-plane wave as a

function of the azimuthal and sagittal angle of the wave vector in the axially symmetric coordinate frame of the head [30,31].

The process of rendering the audible information from an acoustical room simulation to achieve a virtual three-dimensional sound of a given space is generally known as auralization. Virtual acoustics combines the properties of the sound source (directivity and power spectrum), the listener (head-related transfer function HRTF) and the room (binaural room impulse response BRIR) to generate audible information (e.g. auralized sound). The first auralization attempts were monaural and probably date from the era of the first experiments with frequency-transformed music signals. These were played in a scale model by Spandöck in 1934 [32]. It seems that the first computer-based auralization experiments were done in the 1960's by Schröder [33] and in the late 70's by Allen and Barkley [34]. Binaural auralization was introduced a decade later in the work of Pösselt [35], then followed by many others [36-38]. Extensive work on binaural room simulation and scale acoustics models related to spatial hearing was done also by Blauert [39],[40]. In general, virtual modeling of binaural signals in a room is based on convolving anechoic sound samples with the simulated binaural room impulse response (BRIR) and with the transfer function of the loudspeaker. The BRIR-simulation can be obtained in several ways by wave-based or ray-propagation methods. The main wave-based approaches are the finite element method and boundary element method, and ray-propagation methods are based on ray-tracing, cone tracing or image source models [34,41,42].

Wave models are based on solving the wave equations on a 3D grid defined by the user. These models can give accurate results at a single frequency. However, the calculation time is often too long for practical purposes, since a minimum of 6 elements per wavelength is necessary to obtain reliable results. Moreover, the number of natural modes in a room increases with the third power of the frequency and a deterministic approach becomes

redundant, certainly above the limit frequency as pointed out by Schröder [43] already in the 1950's. Therefore, in practical applications finite element methods are used only for the simulation of very small and simple rooms and/or for the lower frequency range. In architectural acoustics, geometrical methods based on the sound propagation along rays are used much more often. These methods usually give reliable results in the high- and middle-frequency range and their calculation time are significantly shorter than any of the wave-based programs [44].

A hybrid calculation method based on the combination of the image source method (ISM) and an adapted ray-tracing method is very popular in architectural acoustics and has provided a good and useful tool for assessing the acoustical quality of the room during architectural design [45,46]. Some work was done on the effects of reverberation in room models on localization [47] as well as on the estimation of localization errors introduced by room reflection [48,49,50]. However, research on validating the suitability of these acoustical models according to sound localization and speech perception for applications in the development and evaluation of new sound processing strategies for auditory prostheses, is non-existing or at least very limited. Therefore, this study aims at validating the hybrid calculation method, by verifying if binaural signals, generated by this method, which involves both a predicted room impulse response and a measured or simulated HRTF, can be used to accurately recreate sound localization cues for a listener in a virtual environment. Although a final judgment about the global auralization quality comprises testing of the accuracy of sound localization for azimuth and elevation angles i.e. testing the front-back confusion when non-individualized HRTFs are use, as well as externalization and overall spectral quality, this paper deals with localization in the frontal horizontal plane, and thus the preservation of binaural ITD and ILD cues, which are addressed in this study, because together with speech reception they are the most relevant for listeners in daily life. For the

validation, a special acoustical environment, i.e. the standardized reverberant room, was chosen. Normal hearing subjects participated in the perception tests.

The global research question of this study can be formulated in the following sub-questions:

(i) What is the difference in localization performance when localizing sound sources naturally or when using headphones? This question comes prior to the other research questions, since ultimate testing of the global simulation tool involves comparing the associated localization performance with the natural situation where listeners do not use headphones.

(ii) What is the difference in localization performance when using headphones and generating stimuli with measured impulse-responses or with virtual acoustics software? This issue is closely related with the accuracy of the hybrid simulation method to predict a BRIR.

(iii) How does reverberation affect localization of sound sources in general? [51,52]

All research questions are answered in detail, including a differentiation in terms of frequency content of the stimuli used. This frequency-dependent analysis helps to distinguish the impact of the different parameters on the binaural cues, as it is well known that the ITD and ILD cues are dominant at respectively low and high frequencies [53-56]. In practice, three different stimuli were used in the experiments: (a) high-frequency noise centered around 3150 Hz, (b) low-frequency sound in the 500 Hz third octave band, and (c) the rather broadband sound of a ringing telephone which also contains a lot of transient information. For the first two sub-questions (i) and (ii), also the influence of the distance between the loudspeaker and the listener, which determines the direct-to-reverberant sound level difference, is investigated [57]. The third sub-question is addressed by comparing the results from listening tests performed in the reverberant room with the ones in anechoic conditions. In an anechoic environment only direct sound reaches the head. In real life, the environment

is such that the sound waves emitted by the source are reflected from different objects or room surfaces, resulting in late reflected wave arrivals from all directions. Therefore, it may be expected that reverberation decreases our ability for directional localization, e.g. the ILD is very sensitive to standing waves in a room, while the ITD quality deteriorates due to the presence of multiple non-coherent reflections. However, as we know from our personal experience, even in very reverberant situations, our ability to localize sound is not completely lost. The influence of reverberation should thus be carefully analyzed.

In section II, the procedures used to obtain the HRTF of the artificial head, and the determination of the BRIR of the reverberant room are elaborated. Their implementation for listening tests is described. The localization performance of the listening subjects in different scenarios is statistically analyzed in section III and discussed in section IV. Conclusions are given in section V.

1 METHODS

1.1 HRTF and BRIR measurements

HRTF measurements on a CORTEX® MK2 manikin artificial head were performed in an anechoic room. The artificial head was placed in the middle of a ring of 2.0 m inner diameter. Thirteen single-cone loudspeakers (FOSTEX® 6301B) of 10-cm diameter were placed every 15° in the frontal plane. A 2-channel sound card (VX POCKET 440 DIGIGRAM®) and DIRAC 3.1 software type 7841 (Bruel and Kjaer Sound and Vibration Measurement Systems) were used to determine for every angle the impulse response for left and right ear by transmitting a logarithmic frequency sweep. Deconvolution of the frequency response of the loudspeaker was not necessary, since for the frequency range of interest (200Hz-10kHz) it was essentially flat (- 3dB points of amplifier response at 20Hz and 50 kHz). The HRTF for left and right ear was determined in the standard way, by calculating the

inverse Fourier transform of the spectrum of the respective recorded sound normalized by the transmitted sweep spectrum (obtained via a hardware loop-back calibration procedure).

These measurements were later used for the creation of sound samples for the listening test in anechoic conditions, by convolving measured dry binaural impulse responses with the three chosen stimuli. The measured HRTF information was also prepared in a plug-in format to be read by the ODEON® software to calculate the BRIRs, which are used to generate the stimuli of the reverberant environment, i.e. to produce the auralization. Since binaural effects in the sagittal plane were not of interest for this research, which is dedicated to localization in the horizontal plane, whatever the sagittal angle of an incoming ray, a value of 0° (i.e. horizontal plane) was assumed for the HRTF. This is equivalent to the assumption that the influence of the sagittal direction on the HRTF is small and independent of the azimuthal dependence.

In the reverberant room, measurements of binaural impulse responses (with the same artificial head as in anechoic room) were done for the same loudspeakers-receiver setup as in the anechoic room (Fig.1, left), i.e. for loudspeakers at 1m distance from the head. In addition, for the investigation of localization of a sound source at larger distances in a reverberant sound field, loudspeakers were put on stands and placed at 2.4 m distance from the artificial head. (Fig.1, right) The 13 measured impulse responses from both experiments were then convolved with anechoic sound samples of the three stimuli for the listening tests with headphones.

Also the impulse response of the headphones placed on the artificial head and recorded via its in-ear microphones was measured using the same sound card and software as for the HRTF characterization above. This transfer function was deconvolved from the sound samples used for listening tests, thus avoiding an unwanted accumulation of the effect of the

transfer function of the ear canal of the artificial head (via the HRTF recording) and that of the ear canal of the listener during the headphone test.

1.2 BRIR Simulations

ODEON® software v.9.0 was used for the room acoustical simulations. This software uses a hybrid algorithm where two geometrical methods are combined to predict the impulse response of a virtual room. The simulation of the room impulse response (RIR) is in principle performed in two steps. The first part, which contains information about early reflections, is calculated by combining the image source method and early scattered rays. The duration of the early part can be chosen by the user via the so-called Transition Order (TO). This is the maximum number of image sources taken into account per initial ray [58]. The second part of the RIR, i.e. the late reflections, is calculated by a modified ray-tracing algorithm that takes into account also the scattering coefficient of the involved surfaces. At every reflection event, local diffuse secondary sources are generated that radiate sound with a directivity according to Lambert's cosine-law [59,60].

This combined approach is quite popular because ISM captures early reflections very accurately, but it is too computationally intensive to calculate all of them for later reflections, so statistical ray tracing is good enough to capture the tail. The hybrid approach has been validated for room acoustics use in a number of Round Robin studies by Vorländer and Bork [61,62].

The Image Source Method is based on the principle that a wave front arriving from a point source and reflecting from an infinite plane can be drawn as if it would originate from an image source. That image source can be evoked as a mirror source considering the reflecting plane in the model as a mirror plane for the incident ray. Also secondary image sources of the initial image sources can be introduced and reflections of the second-order, third-order etc can be calculated. The more surfaces the acoustical model of the room

contains, the more image sources have to be constructed. Obviously, for computational reasons, the exclusive usage of the ISM is suitable only for smaller rooms with a not too complicated geometry [34] and not too much reverberation. In practice, as mentioned before, hybrid methods, combining ISM with RTM are often used.

In the Ray Tracing Method a large number of rays are sent from a point source in a large number of directions, according to the user defined directivity of the loudspeaker or sound source. The trajectory of each ray is determined by its reflection from the boundary surfaces according to Snell's law. The intensity I of a ray decreases with travel distance according to the classical geometrical attenuation of a point source ($I \sim r^{-2}$) and is reduced at every reflection according to the absorption coefficient (defined per octave band) of the incident surface. Scattering of sound is introduced in the computer model via the scattering coefficient δ [%], which is defined as the ratio between the sound energy reflected in non-specular reflections to the total reflected sound energy [59].

The detailed algorithm of the ODEON® software has been described by their developers [63,64]. Essentially, ODEON® software calculates the Binaural Room Impulse Response by transmitting stochastically generated rays from the source that arrive via direct sound or via one or multiple reflections, taking into account appropriate geometrical rules, at the receiver point (in this case, this point is located exactly between both ears). The binaural impulse responses are calculated as follows. First the impulse response is determined for each ray individually when the listener is absent, taking account of the directivity of the sound source and the spectral attenuation of the reflection surfaces encountered on the path from the sound source to the location of the center of the listener's head. Then the left-ear and right-ear HRTFs are used to account for the influence of the listener's head on the incident sound, using the HRTFs corresponding to the direction of the incoming ray (Fig.2). Finally, the responses for the individual rays are summed. The HRTFs used can be input by the user, and

therefore can either be based on measurements on a real subject or an artificial head [30], or computed by a numerical method such as the boundary element method [65]. A geometrical computer model of the reverberant room (Fig.3), which has a volume V = 198 m^3, was based on the measured dimensions. The reverberation time T_{30} [s] of the room was experimentally determined following the ISO 3382 by using an omni-directional point source B&K 4295 and an omni-directional microphone B&K 2642 (Fig.4). During the measurements, equipment (such the 13 loudspeakers) and one test subject were kept in the room, in order to approach the acoustical circumstances of the listening tests. Based on this *in situ* measurement, the acoustical model was calibrated [66].

The loudspeakers (FOSTEX® 6301B) were simulated with the proper directivity and spectrum, as measured in an anechoic room. The impulse response was determined from the signal recorded by a B&K 2642 microphone at 1m from the speaker using the above mentioned DIRAC procedure and Digigram VXPOCKET soundcard and varying the speaker-microphone azimuth angle in steps of 5°. In ODEON® the loudspeaker directivity is introduced by assigning to every stochastically transmitted ray an amplitude that is related to the respective azimutal and elevation angle with respect to the normal to the speaker surface. The receiver properties were defined by the measured HRTF of the artificial head. The simulation was performed with 6000 rays, a maximum reflection order of 2000 and transition order TO = 2, i.e., the image source method was used to calculate the first- and the second-order reflections, and the rest of the RIR was calculated from the special ray tracing method . The influence of the TO value on localization perception tests was investigated by [67]..

1.3 Stimuli

To separately investigate the two binaural cues, ITD and ILD, three different stimuli were presented in the listening tests. As mentioned earlier, for low frequencies (< 1 kHz), the

localization mechanism of a human is dominated by ITD information, while for frequencies above 1.5 kHz, the ILD cue is dominant for directional interpretation [53-56].

To obtain an idea about ITD-based recognition, one-third-octave band noise with central frequency f_c=500 Hz and with duration of 200ms was chosen. Investigation of ILD interpretation was done with one-third-octave band of the same duration (200ms) and a central frequency f_c=3150 Hz. Although 3150 Hz might be considered more upper-middle frequency than really high, thus only just above the region of hair cell phase locking, it is well above 1500 Hz. Our choice of "3150 Hz" sound was guided by a global project goal, which is to test hearing impaired persons, who often suffer from a cut in auditory sensitivity of high frequencies. The third stimulus was a sound of a broadband telephone ring with a duration of 1 second which contains both low- and high-frequency components as well as transient information, and thus was expected to be easier to localize. In order to keep the perception of the telephone sound familiar to the listening subjects, the length of the telephone signal was not shorten to 200ms.

All sound signals were cosine windowed with a rise and fall time of 50ms. The stimuli in all listening tests were presented at 68 dB(A) (Equivalent sound pressure level $L_{A,eq}$ over the duration of the window). A roving level [6] of 4 dB was applied both in tests with and without headphones, i.e. after equalizing the left and right sound pressure levels, they were equally increased with a random variation between 0 dB and 4 dB during the listening test. This procedure allowed listeners to use only ITD and ILD cues, since the sound level of stimuli played from one particular loudspeaker was heard on different sound levels (up to 4 dB difference). This made it impossible to the listening subject to distinguish loudspeakers on the basis of how loud they sound. If the same stimulus was repeated without roving level adjustment several times, the subjects would associate the sound level on which the stimulus was presented with its direction due to the acoustical shadow over the contralateral ear, which

is dependent on the angle of incidence. The brain stem would receive an extra, monaural cue to localize sound. In a real life situation, the absolute level of sound is often unknown, making tests without roving level adjustment not representative

1.4 Subjects

Seven normal hearing subjects between 24 and 50 years old participated in the experiments. They had maximum hearing thresholds of 15dB HL measured at all octave frequencies from 125 Hz to 8 kHz.

1.5 Setups and listening tests

Listening tests were carried out both in the anechoic and reverberant room. The tests were designed in the framework of the research questions mentioned in the introduction. Both rooms were well insulated from outdoor noise, with background noise levels not exceeding 30dB(A). Since the main focus of the research was to investigate the localization ability by using binaural cues, experiments were restricted to the frontal horizontal plane only. Tests of front-back confusion, which relates to monaural spectral cues, were not performed, since this would only complicate the analysis of data.

1.5.1 Anechoic room

For tests in the anechoic room, subjects were seated in the middle of the ring with 13 loudspeakers (labeled by numbers 1-13) located in the frontal horizontal plane every 15° between the azimuth angles -90° and +90° at a distance of 1m, with the subject's ears at the height of the loudspeakers, which were 1.2 m above the floor.

Two listening scenarios were implemented.

(1) Sound samples were played from loudspeakers via a multi-channel soundcard. The listening subject had to identify the loudspeaker in the free field by his or her own ears (OE).

(2) Sound samples based on measured anechoic binaural signals were played from SENNHEISER® HD650 headphones and the listening subject was asked to identify the virtual loudspeaker (HPM).

1.5.2 Reverberant room

Listening scenarios in the reverberant room were created for the real and the simulated reverberant room. In both cases, the listening subjects were sitting in the real reverberant room, identifying the sound emitting loudspeaker:

(1) in situ by his or her own ears (OE)

(2) via headphones based on measurements of head-related transfer functions obtained with the artificial head in the reverberant room (HPM)

(3) via headphones based on the ODEON simulation with the task to identify the virtual loudspeakers in the simulated reverberant room (HPS).

In the environment of the reverberant room, two setups were built (Fig. 1). In setup RR1 the loudspeakers were placed on a ring with inner radius 1m. In setup RR2 the loudspeakers were located on a ring at 2.4m from the listener's head. This was the maximum possible distance in our reverberant room while keeping distance between the loudspeakers and the walls larger than or equal to 1m. The two scenarios represent respectively a situation with a good and one with a poor direct-to-reverberant sound level difference. As a quantity expressing the direct-to-reverberant ratio we have chosen:

$$C_{DR} = 10.\log \frac{\int_{0}^{t_{DR}} p^2(t)dt}{\int_{t_{DR}}^{\infty} p^2(t)dt} \quad [\text{dB}] \qquad (1)$$

where t_{DR} is the early time limit (separating the direct and reverberant sound) and p is the sound pressure. We calculate C_{DR} from the impulse responses measured by the artificial head in the reverberant room placed at respectively 1m and 2.4m from the source. Due to the shadowing of the artificial head and the directivity of the loudspeakers (more directive at higher frequencies), C_{DR} varies with the angle of the arriving sound and increases with frequency. In setup RR1, values for C_{DR} are between -12 and -2dB around 500Hz and between -12 and 8dB around 3kHz. For setup RR2, due to the larger speaker-listener distance, and consequently, the lower direct sound level, the values were lower: C_{DR} varies between -17 and -12dB around 500Hz and between -18 and -1dB for 3 kHz. The negative values of C_{DR} for all angles and both frequencies in setup RR2 indicate that the direct sound is weaker than the reverberant sound.

1.5.3 Listening tests

In all listening tests, the task of the subject was to identify one of the 13 loudspeakers (or an apparent sound source, i.e. a virtual loudspeaker in case of the headphone tests) from which the sound was heard, or the most likely loudspeaker, if sound was recognized e.g. between two loudspeakers. All answers were reported by the listening subject to the operator via a microphone communication system, by announcing the number of the sound source from which the sound was heard. Each subject was asked to continuously keep his/her head pointed towards the central loudspeaker (speaker no.7, at 0° in Fig.1) and was observed by the operator sitting in the next room by using a webcam. Sound samples were played randomly with three repetitions of each speaker in every test. One test therefore consisted of 3x13=39 stimulus presentations or localization tasks. Sufficient time was kept between two tasks, in order to allow the sound of a previous task to sufficiently decay before presenting the next sound. The presentation of the signals and logging of the answers were controlled by

an operator, who controlled the home-made control and data collection program from outside the listening room, while having remote acoustic and webcam contact with the listening subject. For the sake of supplying visual association in the headphone tests (HPM and HPS), the subjects sat in the same position in the room as in the OE experiment, i.e. in the presence of thirteen (inactive) real loudspeakers.. The level of the stimuli in all headphone listening tests was calibrated to 68 dB(A) by means of an artificial ear of B&K®, type 4152. Every subject has participated in the same test two times (test and retest) on two different days, in order to assess the reproducibility of his or her answers. One listening test session took about 45 minutes in anechoic and 90 minutes in reverberant conditions, including one or two breaks. A summary of all performed listening tests can be seen in the Table I.

In the statistical analysis of the results, the root mean squared error (RMS) in degrees is used:

$$\mathrm{RMS}(^\circ) = \sqrt{\frac{\sum_{j=1}^{n}\left(\Theta_{response,j}-\Theta_{true,j}\right)^2}{n}} \qquad (2)$$

where $\Theta_{response}$ and Θ_{true} are respectively the recognized and true azimuthal angle, and n is the number of presented stimuli. In the calculation of the RMS value large errors have a larger impact than small ones. When analyzing the RMS error, the listening subject could choose only from thirteen fixed loudspeaker positions. As a result the smallest error per answer was the distance between two neighbouring loudspeakers, i.e . 15 °. The smallest non-zero error during one test, in which every from thirteen loudspeakers was repeated three times was 2.4 ° RMS.

2 RESULTS AND ANALYSIS

Tables II, III and IV contain the individual and averaged localization results of the 7 normal hearing subjects in the different test conditions. The average values of the test and retest condition are given since no significant difference between both tests are observed. All the data is analyzed using SPSS 15.0. For conciseness, in the following, the term "factorial repeated measures ANOVA" is abbreviated by "ANOVA". All reported pairwise comparisons are Bonferroni corrected for multiple comparisons. The reported p-values are lower bound values and a significance level of $p=0.05$ is used throughout this paper. First a short overview of the presented data is given. Later the ANOVA's, used to examine the research questions, are presented.

In Tables II, III and IV, it is observed that the high-frequency stimulus is localized least accurately. The broadband telephone stimulus is localized best and the data of the 500Hz noise band approaches the data of the telephone signal. Moreover, it is observed that the range of responses, especially those of the anechoic data, corresponds very well with normal hearing data [6]. A second observation is that the localization accuracy of the normal hearing subjects is not drastically influenced by reverberation, nor by how the stimuli are generated or presented to the subjects. This motivates the use of measured or ODEON® generated impulse responses during the first evaluation stages of an algorithm. Statistical analyses are performed to thoroughly evaluate the data. In the first analysis, the difference between natural localization and performing headphone experiments with recorded impulse-responses using a CORTEX MK2 manikin is examined. In the second analysis, the localization performance when using recorded impulse responses is compared to the condition in which impulse responses are generated by ODEON®.

2.1 Natural localization vs. measurements CORTEX MK2 Manikin

An important question related to this paper is whether using impulse-responses measured in an acoustic environment with a CORTEX MK2 manikin has a large influence on localization performance. This is evaluated by examining the own-ear data and the HPM data gathered in the three different acoustic settings, abbreviated as setups AR, RR1 and RR2. The natural localization data, also referred to as own-ear data (OE), and the data of the measured impulse responses (HPM), are inserted in an ANOVA using the following main factors: stimulus type (3150Hz, 500Hz, telephone signal), acoustic environment (AR, RR1 and RR2), stimulus presentation (OE and HPM) and the test-retest factor. First, an interaction between the main factors stimulus type and stimulus presentation is observed ($p=0.004$). Therefore, separate ANOVA's are performed per stimulus. In these three ANOVA's no interactions are observed.

The results are summarized in table V. This table illustrates that a significant decrease in localization performance when using measured impulse responses on an artificial head could be detected only when localizing the 3150Hz stimulus. For this stimulus the OE condition significantly outperforms the HPM condition by, on average, 4.7° RMS. During this analysis no significant effect of reverberation is observed since the performance of AR is not significantly different from RR1 for all stimuli. However, the significant difference between RR1 and RR2, found for both small-band stimuli, do suggest an impact of loudspeaker distance on localization performance. For the telephone signal, no main effect of the acoustic environment was found and hence no pairwise comparisons are calculated.

2.2 Measurements CORTEX MK2 Manikin vs. ODEON software

A second research question motivating these experiments is whether a commercial software package, ODEON, is able to produce virtual sound signals which can be used to perform reliable localization experiments. This is analyzed by using the ODEON data

gathered in the reverberant room. An ANOVA is performed using the main factors stimulus type (3150Hz, 500Hz, telephone signal), acoustic environment (RR1 and RR2), stimulus presentation (OE, HPM and HPS) and the test-retest factor.

Since interactions are found between stimulus type and stimulus presentation, separate ANOVAs are performed for each stimulus. In these ANOVAs interactions are found between stimulus presentation and acoustic environment for the 3150Hz centered noise band and the telephone signal, motivating separate ANOVAs for each acoustic environment for these stimuli. The results of all analyses are presented in table VI. A significant difference in localization performance between the HPM and the HPS condition is only found when localizing the high-frequency stimulus in acoustic environment RR1. In all other experiments, both headphone stimuli give the same localization performance. The other pairwise comparisons are somewhat harder to interpret. It is observed that, in general, headphone experiments seem to reduce localization performance of the 3150Hz stimulus but not of the 500Hz stimulus which was also concluded in one of the previous paragraphs (OE vs. HPM and HPS). When localizing a telephone stimulus this trend was observed in the condition RR2.

3 DISCUSSION

The main purpose of the performed tests is related to comparing listening performance between simulated sound presented via headphones on one hand and natural hearing on the other hand, as well as assessing the quality of simulated BRIRs in comparison with measured BRIRs. Before going into those aspects, we draw attention to the following observation. Regardless of the circumstances, the high-frequency stimulus is localized least accurately. For the subjects under study, the ILD cue, which is predominantly present in the higher part of the frequency spectrum, was apparently processed less adequately than low-

frequency ITD information. In the discussion below, the possible reasons for deterioration by specific circumstances should therefore be interpreted mainly via their impact on the ITD cue, which is expected to dominate the subjective localization assessment. Visual feedback during the listening tests might influence the result as well. E.g., due to the visual feedback and the presence of loudspeakers in the frontal plane only, test subjects did not report on perception of sounds coming from above or behind the head.

3.1 Difference in localization performance when localizing sound sources naturally or when using headphones

To address this issue, comparison of the listening scenarios HPM and HPS is necessary, since they use the same HRTF information. In the condition HPM, impulse responses measured by the artificial head were performed in the acoustic environment of the localization experiment. Afterwards, these impulse responses were used for the headphone stimuli generation. In the condition HPS, a virtual acoustics software was used to generate the reverberant stimuli used in the headphone experiment. This included the combination of measuring HRTFs of an artificial head in an anechoic environment and modelling the room in the ODEON® software. Table V summarizes the statistical analysis done on the data of the anechoic room (AR) and both settings in the reverberant room (RR1 and RR2).

This table indicates that the low-frequency noise is localized equally well using headphones as in real life. This is confirmed by the data in Table VI which shows that, for this stimulus, both headphone conditions have a similar performance compared to the natural localization. This suggests that the ITD cues measured between the microphones of the CORTEX MK2 and which is determined by the positioning of the ear simulators in the manikin, sufficiently approaches the ITD cues normally used by the subjects.

The data for the high-frequency noise component on the other hand shows that, in general, a significant decrease in localization performance is present when localizing sounds under headphones compared to using own ears. This is observed in Table V and in 3 out of 4 pairwise comparisons in Table VI. When analyzing the intra-subject differences in Tables II, III and IV, decreases in performance in the order of 8° to 9° are often observed with a maximum decrease of 14.2° (S6, Table III). Localization of the 3150Hz centered noise band is mainly based on using ILD information which is introduced by diffraction and reflections of sounds around and on the head and torso of a human listener. The observed significant decrease in localization performance may be explained by differences in acoustical properties between an artificial head and a human listener. These differences are due to differences in shape, in material (the artificial head is made out of a hard synthetic material) and due to a lack of clothing and hair which all have a significant impact on ILD cues.

For a broadband telephone signal, it is observed that when taking all environments (Table V) into account, no significant decrease in localization performance is found. When isolating the data of the reverberant environment (Table VI) a significant difference is observed, but only if the sound sources are placed at 2.4m distance from the subject. However, when analyzing the differences made in this condition (Table IV), it is observed that the intra-subject differences between conditions OE and HPM are in range of -0.5° to 3.0°. The intra-subject differences between OE and HPS are in the range of 1.5° to 5.2° RMS except for subject S6 which showed a decrease of 14.0° RMS. These differences may be regarded as acceptable depending on the experiment.

It can be concluded that significant differences can be present between naturally localizing sound sources or localizing sound sources under headphones. However, these differences are mainly present when localizing narrow-band high-frequency stimuli. When using lower frequencies or broadband stimuli, no or smaller differences have been observed.

The observed differences seem to be due to the artificial head. This conclusion is supported by two studies of Møller *et al.* In the first study [68] no significant difference in localization performance was found between natural localization and headphone experiments if individual recording were used. In the second study [69], a significant decrease in localization performance was observed when using recordings of 8 different artificial heads (not including the CORTEX MK2) in the same localization setup. By isolating the so-called out-of-cone errors, which are related to the horizontal localization performance studied here, 7 out of 8 artificial heads introduced a decreased localization performance. A second evaluation, in the same localization setup, by Minnaar *et al* [70], using more recent artificial heads, demonstrated that artificial head recordings were improving and there is reason for optimism concerning an evolution towards artificial heads approaching real-life performance better and better.

3.2 Difference in localization performance when generating stimuli with measured impulse-responses or with virtual acoustics software

Evaluations were performed using HPM and HPS stimuli in a reverberant room. Table VI indicates that a significant difference between HPM and HPS is observed only when localizing high-frequency sound sources which are positioned at 1m from the subject. The errors introduced by the ODEON® software for this specific condition range from 0.7 to 11.3°. In contrast with condition RR1, ODEON® does not introduce a large decreases in localization performance in condition RR2. Moreover, in this condition, performance even improves for 5 out of 7 subjects. This seems contradictory. However, when examining the HPM data, which is the reference condition for this research question, of the 3150Hz noise band, it is observed that the RR1 condition significantly outperforms the RR2 condition. By using the ODEON® software the performance of RR1 is decreased to the level of RR2. In

other words, ODEON® somewhat affects the localization performance of a high-frequency stimulus if a high performance is obtained at the baseline condition. When using low and broadband stimuli no significant differences are observed (Table VI).

It can be concluded that ODEON® software introduces some binaural cue distortion compared to HPM stimuli when using high-frequency signals. Possibly this is related to the extrapolation of the horizontal plane HRTF used to all sagittal angles. These distortions lead, to a certain extent, to a significant decrease in localization performance. If the originally obtained localization performance by using HPM stimuli is moderate, ODEON® will not introduce additional distortions.

3.3 Influence of reverberation on the natural localization of sound sources or on the localization of sound sources under headphones

During these experiments, two different extreme acoustic environments have been studied: the anechoic room and the reverberant room. By observing the data in Table V it is clear that, for all stimuli, reverberation time as such does not influence localization performance (AR vs RR1) when loudspeakers are sufficiently close to the subject. However, interestingly a significant difference is observed when changing the distance between the loudspeakers and the subject (RR1 vs RR2). The fact that the reverberation does not affect localization performance can be explained by the so called precedence effect which is also known as the law of the first wavefront (for an overview on the precedence effect see [71]). The precedence effect is based on the ability of the human auditory system to associate the direction of arrival of a sound source with the direction of arrival of the direct sound. The time interval in which this takes place is called the fusion zone. Outside the fusion zone, reflecting sounds are perceived as being echoes which have their own direction of arrival.

These evaluations show that the precedence effect avoids significant differences in localization performance between two very extreme acoustic environments.

The significant difference in performance between condition RR1 and RR2 is less straightforward to explain. Two interpretations are possible. The first explanation is that localizing sound sources positioned at 2.4 m distance is less accurate than localizing sound sources at 1m distance in general. However, this cannot be concluded from the data, since no evaluations were done in a 2.4 m setup in the anechoic environment. Moreover, in the work of [72], the claim was made that ITD and ILD cues are virtually independent of distance, if the sound sources are positioned beyond 1 m. The difference in localization performance between scenario RR 1 and RR2 can be explained in an alternative way by inferring that due to the lower direct-to-reverberant ratio, the first wavefront in RR2 is too weak with respect to the subsequent sound from reflections to allow for a correct sound localization of the sound source [57]. This suggests that placing the sound sources further would affect the localization performance even more. However, this was not investigated, because of the limited room dimensions. We also limited the stimuli in this study sounds with a fairly transient character. One can expect that localization on the basis of continuous stimuli would suffer more in a reverberant situation.

4 CONCLUSION

A comparison was made between localizing sounds in a natural way and localizing sound sources in simulated binaural sound presented via headphones. Two different headphone conditions were evaluated. First, impulse-responses were measured between loudspeakers and an artificial head, and subsequently used to generate the headphone stimuli. A second set was generated by using ODEON® software enabling the acoustic modeling of virtual environments.

Presenting listening stimuli via headphones tends to decrease the localization performance only and slightly when localizing narrow-band high-frequency stimuli. The use of an artificial head for generating the HPM and HPS stimuli may explain these differences, since the acoustical properties of an artificial head are different from the ones of a human listener, thereby generating less natural ILD cues. Also, for headphone tests, neglecting the sagittal angular dependence of the measured HRTF could have a slightly deteriorating effect. When localizing narrow-band low-frequency signals or a broadband telephone signal, either no or a small decrease in performance was observed, since geometrical head details are less relevant for long wavelengths.

In most of the virtual scenarios, the localization performance is almost as good as the one observed for the natural scenario. Headphone experiments show little decrease in localization performance compared to measured impulse-responses in only one out of six test conditions, namely when using narrow-band high-frequency signals.

Reverberation has no significant influence on localization accuracy if the distance between the receiver and a loudspeaker is only 1 m. This may be understood in view of the the preserved dominance of the direct sound helping the listener to localize the source. However, in the same room, for a larger distance (2.4 m) between loudspeakers and subject, the localization performance deteriorates. This indicates that the impact on the localization performance of a decreasing direct-to-reverberant sound ratio [57] is higher than the impact of reverberation as such. The result also shows that sound source localization is more robust under reverberant conditions than speech recognition, which typically rapidly deteriorates in the presence of reverberation, due to the highly energetic noise background resulting from long persisting echoes. Even a moderate deterioration of sound source localization performance can thus be seen as an indicator for drastic reduction of speech intelligibility,

partly due to the background noise, and partly due to the loss of effective S/N enhancement that is brought by binaural sound reception.

The good overall agreement between the simulation- and measurement-based results in localization tests proves that the simulation approach does a good job in terms of predicting BRIRs, useful for directional-localization perception test in the frontal horizontal plane. In support of this conclusion, it is worth mentioning that in a related article [73] it was shown that using predicted BRIRs for sound source localization tests in a typical virtual classroom, angular asymmetry was revealed in the localization performance, which was due to the asymmetric location of sources and listening subject with respect to the objects in the classroom.

Future research should determine if localization performance differences at high frequencies persist when the HRTF of the artificial head is replaced by an HRTF that is individually recorded by placing intra-ear microphones in individual listening subjects, or by the HRTF obtained from finite element models of artificial and test subject's heads.

This research is also a first step towards verifying whether using simulated BRIRs in listening tests in virtual acoustic scenarios are adequate for speech recognition assessment, both for normal hearing persons and for hearing impaired persons wearing hearing aids with cue processing algorithms to be tuned for real life circumstances.

5 ACKNOWLEDGEMENTS

We would like to thank the developers of the ODEON® software, in particular Prof. Jens-Holger Rindel and Claus Lynge Christensen for their help and advice while working with the ODEON® software. We also thank all the people who participated in the listening tests.

This research was financed by FWO-V project G.0334.06.

6 TABLES

TABLE I : A summary of all listening tests. Setup AR is for the anechoic room, Setup RR1 is for the reverberant room where the distance between the loudspeakers and receiver was 1 m, Setup RR2 is for the reverberant room where the distance between the loudspeakers and receiver was 2.4 m

TABLE II : The individual and average RMS data (°) of the normal hearing subjects for the different test conditions. Tests are done in an anechoic room with a distance between the loudspeakers and the subject of 1m (AR)

TABLE III : The individual and average RMS data (°) of the normal hearing subjects for the different test conditions. Tests are done in the reverberant room with a distance between the loudspeakers and the subject of 1m (RR1)

TABLE IV : The individual and average RMS data (°) of the normal hearing subjects for the different test conditions. Tests are done in the reverberant room with a distance between the loudspeakers and the subject of 2.4m (RR2)

TABLE V : p-values of the pairwise comparisons using the data OE and HPM for the three different acoustic settings. A significant difference between OE and HPM is only observed for the 3150Hz centered noise band. The term "no" indicates that no main effect was found, hence no pairwise comparisons are performed.

TABLE VI : p-values of the pairwise comparisons using the data OE, HPM and HPS in the reverberant room. A significant difference between HPM and HPS is only observed when

localizing the 3150Hz cantered noise band arriving from 1m distance of the subject. The term "no" indicates that no main effect was found, hence no pairwise comparisons are performed.

TABLE I

	Anechoic room						Reverberant room														
Set up's	Set up AR						Set up RR1									Set up RR2					
Listening scenarios	OE			HPM			OE			HPM			HPS			OE			HPM			HPS		
Stimuli	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone	500Hz	3150Hz	telephone

TABLE II

RMS error (°)	500 Hz		3150 Hz		telephone	
	OE	HPM	OE	HPM	OE	HPM
S1	12.9	12.9	19.3	17.9	8.0	6.8
S2	10.6	9.5	16.7	22.6	6.4	6.8
S3	12.1	10.8	17.0	21.8	6.6	9.9
S4	8.2	8.8	13.9	16.6	6.2	7.0
S5	8.4	14.8	16.0	20.7	7.2	9.8
S6	15.2	14.6	18.8	23.6	10.3	10.9
S7	9.0	8.3	18.9	24.7	6.8	3.3
average	10.9	11.4	17.2	21.1	7.3	7.8
stdev	2.6	2.7	1.9	2.9	1.4	2.6

TABLE III

RMS error (°)	500 Hz			3150 Hz			telephone		
	OE	HPM	HPS	OE	HPM	HPS	OE	HPM	HPS
S1	11.6	12.1	15.0	13.1	13.9	20.0	9.0	10.8	11.6
S2	10.4	10.0	9.8	17.1	22.9	27.6	8.1	5.6	6.9
S3	11.1	13.1	9.2	19.4	18.5	26.9	8.2	9.6	7.6
S4	7.6	8.1	10.0	14.3	16.0	17.5	5.5	7.0	9.3
S5	10.4	10.6	12.1	12.2	13.1	24.4	7.2	9.0	10.9
S6	15.1	18.4	18.2	11.1	18.4	25.3	11.0	10.3	17.4
S7	11.6	11.2	10.4	20.3	26.1	26.8	9.5	11.5	10.0
average	11.1	11.9	12.1	15.3	18.4	24.0	8.3	9.1	10.5
stdev	2.2	3.3	3.3	3.6	4.7	3.9	1.7	2.1	3.5

TABLE IV

RMS	500 Hz			3150 Hz			telephone		
error (°)	OE	HPM	HPS	OE	HPM	HPS	OE	HPM	HPS
S1	14.5	13.6	11.8	19.1	28.5	20.1	9.0	8.5	11.7
S2	13.6	12.6	12.5	18.0	27.1	29.9	9.0	9.6	13.2
S3	13.2	13.6	13.5	17.6	25.9	22.3	9.2	10.6	14.0
S4	13.0	10.6	11.9	17.0	26.0	21.1	7.3	10.2	11.7
S5	11.9	11.6	10.6	17.8	22.4	19.1	8.6	10.8	13.8
S6	21.4	21.7	26.0	23.5	25.6	30.5	10.1	13.1	24.1
S7	18.9	17.1	16.7	17.7	25.9	25.4	16.1	17.6	17.6
average	15.2	14.4	14.7	18.7	25.9	24.0	9.9	11.5	15.1
stdev	3.5	3.8	5.3	2.2	1.9	4.6	2.9	3.0	4.4

TABLE V

	500 Hz	3150 Hz	Telephone
AR vs RR1	1.000	0.354	no
AR vs RR2	0.074	0.022*	no
RR1 vs RR2	0.014*	0.047*	no
OE vs HPM	0.744	<0.001*	0.067

TABLE VI

	500Hz	3150Hz		Telephone	
		RR1	RR2	RR1	RR2
RR1 vs RR2	0.015*	-	-	-	-
OE vs HPM	no	0.126	0.001*	no	0.044*
OE vs HPS	no	0.003*	0.030*	no	0.042*
HPM vs HPS	no	0.021*	0.966	no	0.094

7 FIGURE CAPTIONS

Figure 1: Left: setup in the anechoic room. Right: setup in the reverberant room. An array of 13 speakers is positioned in the frontal horizontal hemisphere of the subject or manikin. The speakers are placed at 1m of the subject. In the reverberant room, a second array of loudspeakers is used with a radius of 2.4m. In all setups the impulse responses between the loudspeakers and a CORTEX MK2 manikin are measured. These impulse responses are used to generate stimuli which are presented over headphones. The impulse responses measured in the anechoic room are also used in ODEON® to generate a second set of headphone stimuli.

Figure 2: Illustration of the BRIR calculation

Figure 3: Geometrical model of the reverberant room in the ODEON® software

Figure 4: Reverberation time T_{30} [s] as measured in the reverberant room according to ISO 3382 (2004).

Figure 5: Summary of the mean localization performance of the 7 normal hearing subjects together with their inter-subject standard deviation. (The full data sets are given in Tables II to IV.) A '*' depicts a significant difference with the OE condition. A 'x' depicts a significant difference in performance between both headphone conditions, HPM and HPS. Significant differences are mainly present when localizing high-frequent stimuli. Overall, the differences between conditions are fairly small, especially when localizing broadband or low-frequency stimuli.

Reference List

[1] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient-based implementation of Spatially Preprocessed Speech Distortion Weighted Multichannel Wiener filtering for noise reduction in hearing aids," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 911-925, (2005).

[2] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. van Dijk, A. van Wieringen, and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM (TM) in the nucleus Freedom (TM) cochlear implant system," *Ear and Hearing*, vol. 28, no. 1, pp. 62-72, (2007).

[3] J. B. Maj, L. Royackers, J. Wouters, and M. Moonen, "Comparison of adaptive noise reduction algorithms in dual microphone hearing aids," *Speech Communication*, vol. 48, no. 8, pp. 957-970, (2006).

[4] H. Dillon, *Hearing Aids*, Boomerang Press, Sydney, (2001).

[5] A. W. Bronkhorst , "The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica United with Acustica*, vol. 86, pp. 117-128, (2000).

[6] T. Van den Bogaert, T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters, "Horizontal localization with bilateral hearing aids: Without is better than with," *Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 515-526, (2006).

[7] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 484-497, (2008).

[8] ANSI S 3-5:1997, "Methods for calculation of the Speech Intelligibility Index, American National Standard S3.5-1969, Acoustical Society of America," (1997).

[9] IEC 60268-16:2003, *Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, (2003).

[10] H. J. M. Steeneken, and T. Houtgast, "Physical Method for Measuring Speech-Transmission Quality," *Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 318-326, (1980).

[11] T. Brand, and R. Beutelmann, "Examination of an EC/SII based model predicting speech reception thresholds of hearing-impaired listeners in special noise situations," *In Proceedings of the 21st Danavox symposium "Hearing Aid Fitting"*, pp. 139-151, (2005).

[12] R. Beutelmann, and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331-342, (2006).

[13] www.odeon.dk, last viewed on 10th February 2009.

[14] W. Yang, and M. Hodgson, "Validation of the auralization technique: Comparative speech-intelligibility tests in real and virtual classrooms," *Acta Acustica United with Acustica*, vol. 93, no. 6, pp. 991-999, (2007).

[15] G. Keidser, K. Rohrseitz, H. Dillon, V. Hamacher, L. Carter, U. Rass, and E. Convery, "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers," *International Journal of Audiology*, vol. 45, no. 10, pp. 563-579, (2006).

[16] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *Journal of the Audio Engineering Society*, vol. 9, pp. 148-151, (1961).

[17] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. Patent 3,236,949, Feb. 1966 (filed 1962) cited in Y. Huang, J. Benesty, and J. Chen, "On Crosstalk Cancellation and Equalization With Multiple Loudspeakers for 3-D Sound Reproduction," *IEEE Signal processing letters*, vol. 14, no. 10, October 2007.

[18] H. Møller, "Reproduction of Artificial-Head Recordings Through Loudspeakers," *Journal of the Audio Engineering Society*, vol. 37, no. 1-2, pp. 30-33, (1989).

[19] D. B. Ward, and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," *IEEE Signal Processing Letters*, vol. 6, no. 5, pp. 106-108, (1999).

[20] D. Ward, and G. Elko, "A robustness analysis of 3D audio using loudspeakers," *In Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99*, pp. 191-194, (1999).

[21] M. S. R. Bai, and C. C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on subband filtering," *Journal of Sound and Vibration*, vol. 290, no. 3-5, pp. 1269-1289, (2006).

[22] M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1056-1069, (2007).

[23] P. Damaske, "Head-Related 2-Channel Stereophony with Loudspeaker Reproduction," *Journal of the Acoustical Society of America*, vol. 50, no. 4, pp. 1109-&, (1971).

[24] K. C. K. Foo , M. O. J. Hawksford, and M. P. Hollier , "Optimization of sound reproduced using two loudspeakers," *In Proceedings of the 16th AES International Conference: Spatial Sound Reproduction*, (1999).

[25] M. R. Bai, C. W. Tung, and C. C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 2802-2813, (2005).

[26] T. Takeuchi, and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786-2797, (2002).

[27] J. Merimaa, and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115-1127, (2005).

[28] V. Pulkki, and J. Merimaa, "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," *Journal of the Audio Engineering Society*, vol. 54, no. 12, pp. 3-20, (2006).

[29] D. Hammershoi, and H. Moller, "Methods for binaural recording and reproduction," *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 303-311, (2002).

[30] H. Møller, M. F. Sorensen, D. Hammershoi, and C. B. Jensen, "Head-Related Transfer-Functions of Human-Subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300-321, (1995).

[31] F. L. Wightman, and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648-1661, (1992).

[32] F. Spandöck, "Raumakustische Modellversuche (Roomacoustical simulations)," *Annalen der Physik*, vol. 20, no. 4, pp. 345-360, (1934).

[33] M R. Schroeder, C. Atal, and B. S. Bird, "Digital Computers in room acoustics," *In Proceedings of the 4th Int.Cong.Acoustics, Copenhagen, Denmark*, (1962).

[34] J. B. Allen, and D. A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, (1979).

[35] C. Pösselt, "Binaurale Raumsimulation für Kopfhörerwiedergabe (Binaural simulation for reproduction via headphones)," *In Proceedings of DAGA '87*, (1987).

[36] K. H. Kuttruff, "Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 876-880, (1993).

[37] M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization - An Overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861-875, (1993).

[38] M. Vorländer, *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, RWTH edition, Springer Verlag, (2008).

[39] J. Blauert, H. Lehnert, W. Pompetzki, and N. Xiang, "Binaural Room Simulation," *Acustica*, vol. 72, no. 4, pp. 295-296, (1990).

[40] N. Xiang, and J. Blauert, "Binaural Scale Modeling for Auralisation and Prediction of Acoustics in Auditoriums," *Applied Acoustics*, vol. 38, no. 2-4, pp. 267-290, (1993).

[41] T. Lewers, "A Combined Beam Tracing and Radiant Exchange Computer-Model of Room Acoustics," *Applied Acoustics*, vol. 38, pp. 161-178, (1993).

[42] I. A. Drumm, and Y. W. Lam, "The adaptive beam-tracing algorithm," *Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1405-1412, (2000).

[43] M. Schröder, "Die statistischen Parameter der Frequenzkurven von grossen Räumen (The statistical parameter of the frequency responses of large rooms)," *Acustica*, vol. 4, pp. 594-600, (1954).

[44] X. Zeng, C. Christensen, and J. Rindel, "Method of modeling sound scattering in rooms at low frequencies," *Acta Acustica*, vol. 31, no. 5, pp. 476-480, (2006).

[45] J. H. Rindel, and C. L. Christensen., "Applications of a system for two-channel auralization using Odeon-Sound," *In Proceedings of the 3rd Join Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu*, (1996).

[46] M. Rychtáriková, L. Nijs, and G. Vermeir, "Auralization as a Tool to Predict the Acoustical Quality of Open Plan Offices," *In Proceedings of the Internoise 2005, the 34th International Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brasil*, (2005).

[47] D. R. Begault, and E. M. Wenzel, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904-916, (2001).

[48] W. M. Hartmann, "Localization of Sound in Rooms," *Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1380-1391, (1983).

[49] B. Rakerd, and W. M. Hartmann, "Localization of Sound in Rooms .2. the Effects of A Single Reflecting Surface," *Journal of the Acoustical Society of America*, vol. 78, no. 2, pp. 524-533, (1985).

[50] W.M.Hartmann. "Auditory localization in rooms," *In Proceedings of AES 12th International Conference*, pp. 34-39, Copenhagen, Denmark (1993).

[51] N. Kopco, and B. Shinn-Cunningham, "Auditory Localization in Rooms: Acoustic Analysis and Behavior," *In Proceedings of the 32nd International Acoustical Conference - EAA Symposium "Acoustics Banská Štiavnica, September*, (2002).

[52] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse response," *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3100-3115, (2005).

[53] A. W. Bronkhorst, "Localization of Real and Virtual Sound Sources," *Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542-2553, (1995).

[54] W. M. Hartmann, "How We Localize Sound," *Physics Today*, pp. 24-29, (1999).

[55] L. R. Bernstein, and C. Trahiotis, "Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli"," *Journal of the Acoustical Society of America*, vol. 112, no. 3, pp. 1026-1036, (2002).

[56] L. R. Bernstein, C. Trahiotis, and R. L. Freyman, "Binaural detection of 500-Hz tones in broadband and in narrowband masking noise: Effects of signal/masker duration and forward masking fringes," *Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2981-2993, (2006).

[57] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2110-2117, (2002).

[58] M. Lisa, J-H Rindel, and C. L Christensen, "Predicting the Acoustics of Ancient Open-Air Theatres: The Importance of Calculation Methods and Geometrical Details," *In Proceedings of the Joint Baltic-Nordic Acoustics Meeting*, (2004).

[59] C. L. Christensen, and J. Rindel, "A new scattering method that combines roughness and diffraction effects," *In Proceedings of Forum Acusticum, Budapest, Hungary,* (2005).

[60] X. Zeng, C. L. Christensen, and J. H. Rindel, "Practical methods to define scattering coefficients in a room acoustics computer model," *Applied Acoustics*, vol. 67, no. 8, pp. 771-786, (2006).

[61] M. Vorländer, "International round Robin on room acoustical computer simulations," *In Proceedings of the 15th International Congress on Acoustics (ICA '95), Trondheim, Norway, June*, (1995).

[62] I. Bork, "Report on the 3rd Round Robin on Room Acoustical Computer Simulation - Part II: Calculations," *Acta Acustica United with Acustica*, vol. 91, no. 4, pp. 753-763, (2005).

[63] J. H. Rindel, "The Use of Computer Modeling in Room Acoustics," *Journal of Vibroengineering*, vol. 2, no. 4, pp. 219-224, (2000).

[64] C. L. Christensen, *Odeon Room Acoustics Program, Version 9.0, User Manual, Industrial, Auditorium and Combined Editions*, Odeon A/S, Lyngby, Denmark, (2007).

[65] J. Fels, P. Buthmann, and M. Vorlander, "Head-related transfer functions of children," *Acta Acustica United with Acustica*, vol. 90, no. 5, pp. 918-927, (2004).

[66] J. H. Rindel, and C. L Christensen, "Room acoustic Simulation and Auralization – How close can we get to the real room?," *In Proceedings of WESPAC 8, The Eighth Western Pacific Acoustics Conference, Melbourne, 7-9 April 2003,* (2003).

[67] M. Rychtáriková, and G. Vermeir, "Comparison Between Image Source Method and Ray tracing method in the Binaural Room Impulse Responses," *In Proceedings of the 2nd International Symposium "MAP '06", Zvolen, Slovakia, September,* (2006).

[68] H. Møller, M. F. Sorensen, C. B. Jensen, and D. Hammershoi, "Binaural technique: Do we need individual recordings?," *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451-469, (1996).

[69] H. Møller, D. Hammershoi, C. B. Jensen, and M. F. Sorensen, "Evaluation of artificial heads in listening tests," *Journal of the Audio Engineering Society*, vol. 47, no. 3, pp. 83-100, (1999).

[70] P. Minnaar, S. K. Olesen, F. Christensen, and H. Moller, "Localization with binaural recordings from artificial and human heads," *Journal of the Audio Engineering Society*, vol. 49, no. 5, pp. 323-336, (2001).

[71] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633-1654, (1999).

[72] D. S. Brungart, and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465-1479, (1999).

[73] M. Rychtáriková, J. Fels, G. Vermeir, J. Wouters, and M Vorländer, "Localizing The Teacher In The Virtual Classroom," *In Proceedings of the 19th International Congress On Acoustics - ICA, Madrid, 2-7 September*, (2007).